

Alena Anishchanka, Dirk Speelman, Dirk Geeraerts
University of Leuven, Belgium

SOCIOLECTAL VARIATION IN THE LENGTH OF COLOR TERMS USED IN ADVERTISING

Abstract

In the article we explore the length of color terms used in advertising as one of the formal parameters of their lexical complexity. We report two case studies applying multiple linear regression to test the variation in the color term length relative to two sets of sociolinguistic parameters relevant for online marketing and purchase. The first case study reveals the most general variation patterns in four product categories (cars, clothing, women's makeup, house paints) involving the effects of such sociolinguistic factors as product category and the status of the brand. In the second case study, we propose a more fine-grained analysis of the product category of cars focusing on the product-specific sociolinguistic parameters (type of car, year of production, country of the parent company) and their interactions with the more general variables revealed by the first analysis.

Key words: color term length, sociolectal variation, color categorization, advertising, multivariate linear regression method

1. Introduction

Probably one of the first things that 'catches the eye' of a linguist or a layperson when faced with the color terms used in advertising is their sheer diversity and complexity. Ranging from quite innocent *burgundy*, *plum* and *orchid* to *Who the Shrek are you?* or *Teal the cows come home*, these names stretch the creative potential of a language to the extreme where they can hardly be recognized as color terms any more. Such words reveal the richness of associative meanings and metaphorical extensions

in color vocabulary that has been successfully exploited for commercial purposes and has served as an inspiration to artists like Peter van Eyck. In his exhibition “Elusive Colors”, inspired by color names in car catalogs, he compared color names such as *whisper green* or *virtual blue* to ‘industrial poetry’.

Going beyond the purely entertaining aspects of color terms, linguists have increasingly recognized advertising as an important cultural factor in the development of color vocabulary and color categorization (see, for instance, Kertulla 2002: 276-277). With advertising becoming an ever more present part of everyday life, color terms and color categories that were once exotic, rare or specialized, such as *teal*, *indigo*, *khaki*, or *vanilla* become more and more present in everyday purchasing situations. Thus the development of online marketing, advertising, and image technology can be seen as yet another step in the evolution of technology, whose impact on color categorization has been demonstrated in a number of studies (Casson 1992; MacLaury 1992, 1997).

At the same time the multiplicity and complexity of color terms in advertising, their arbitrary and open-ended character make analyzing them a challenging and even intimidating task. In her study of fashion color terms, Stoeva-Holm questions “if certain color terms which are used for fashion escape classification altogether, possibly on account of their being neologisms or ad-hoc formations?” (1996: 423). These characteristics of color terms set a number of constraints on the studies of color vocabulary in modern advertising, especially those using an introspective method. Firstly, the number of color terms included in the analyses remains quite limited, not exceeding several hundred tokens. For comparison, the online color naming experiment carried out by HP Lab yielded over 30 thousand color words, demonstrating the color naming potential of the English language (Moroney 2008). Secondly, introspective methodology limits the scope of the analysis in terms of covered domains. The studies known to us are limited to color names in one or at best two product categories, such as cars (Bergh 2007; Graumann 2007), fashion (Klaus 1989; Cencig 1990; Stoeva-Holm 1996, 2007), textiles and cosmetics (Whyler 2007). As a result, there always remains the question of how general the proposed generalizations and conclusions are and to what extent they are applicable to related domains, product categories or genres. These constraints in many ways also apply to corpus-based studies of color vocabulary, mostly due to the limited size of the existing corpora, which do not allow the

analysis of more than a few hundred most commonly used color words (Steinvall 2002). For instance, Kertulla points at the limitations of the traditional corpus / dictionary resources (such as OED) for studying color terminology when it comes to the 20th century, since the current color names are not adequately reflected, especially their current usage and development (2002: 273).

Considering all that, in our study we specifically address the issue of color term complexity and variation in advertising, which raises a number of questions. When we say color terms are complex, do we mean they have complex morpho-syntactic structure, like *dream a little dream of me* or *obsidian black pearl onyx metallic*? Do we mean they are complex in the subtlety of the conveyed color nuances, like *nocturne aubergine*, *creamy blueberry* or *baby's breath*? Or do we mean they are complex in the associative meanings they evoke, ranging from *pine apple upside down cake* to *ms.behave* or *fusion orange*? Narrowing down our research question, in this study we focus on one of the formal aspects in the complexity of color names, specifically, on the length of color terms, and analyze its variation relative to sociolectal factors relevant for an online purchasing situation. Other aspects of complexity in color names we intend to explore in separate studies.

The length of color terms and their frequency have been traditionally used as standard measurements in the studies of basic color terms and categories in the Berlin and Kay tradition (for a summary see Corbett and Davies 1997). Taking a typological perspective, these studies investigate the varying length of basic color terms across languages, using this parameter as one of the basicness criteria. In line with Zipf's law, the general argument suggests that basic color terms are more frequently used and are shorter than other color words. In our case studies, we address language-internal variation and take this measurement (far) beyond the basic color terms. We will try to demonstrate that the length of color terms as a formal parameter of lexical complexity is subject to variation across advertised products and is sensitive to contextual factors such as product category, brand status or country of parent company. This kind of categorization study continues the lines of research set in Geeraerts, Grondolaers, Bakema (1994) that explored different types of lexical variation using the domain of clothing as a starting point.

In our analyses, we seek to contribute to the study of color naming and categorization in two ways. At the theoretical level, we take a cognitive sociolinguistic perspective on color naming (Kristiansen, Dirven 2008;

Geeraerts, Kristiansen, Peirsman 2010). This implies that we take a usage-based approach in line with current tendencies of contemporary lexical semantics as described in Geeraerts (2010) and analyze color categories in context, taking into account the multiple parameters of a specific discourse situation. This approach requires the use of real-language data, in this case an extensive self-constructed corpus of online marketing materials including color words, color samples and sociolectal information present in online advertising.

At the methodological level, we propose a quantitative corpus-driven analysis based on multivariate (regression) statistical techniques. Our aim here is to develop a rigorous methodology that would allow testing the insights made in smaller scale descriptive studies of color names in advertising. Thus the chosen granularity of our analyses allows bringing together the perspectives from coarse-grained basic color categorization studies and very fine-grained studies of color naming in advertising (Bergh 2007; Graumann 2007, Cencig 1990, Klaus 1989; Stoeva-Holm 1996, 2007; Whyler 2007). We believe that this combined approach can give a better understanding of the lexicogenetic tendencies in color naming in modern advertising.

2. The study design

In the following sections, we report the results of two case studies exploring the effects of product-related sociolectal factors on the length of the color terms used in American advertising. The two analyses use the same method of multivariate linear regression, but are different in their scope. In the first study we explore the variation in the length of color terms across four product categories (cars, clothes, makeup, paints), and in the second analysis we zoom in on the product category of cars to include more product-specific factors. This allows us to look at different sets of sociolectal factors involved in linguistic variation.

We start with presenting a somewhat unusual type of corpus used as the material for the analyses. In section 3 we report on the first case study. After presenting the response variable and sociolectal parameters included as explanatory variable in the regression models, we discuss the results of the regression analysis measuring the length of color terms used in four product categories relative to the most general contextual factors,

such as product category, the prestige status of the brand, and diversity in product colors. In section 4, we report the second case study zooming in on the product category of cars, and focusing on additional product-related factors such as the country associated with the parent company, year of production, type of car. In the final section we compare the results of the analyses and suggest possible future extensions of the study.

2.1 Data

Our analyses are based on samples from an extensive self-constructed corpus including 70620 tokens of color term usage in online marketing materials. The data were collected manually and automatically (using Python scripts¹) from the websites developed by manufacturing and retail companies for advertising and sales in the US market. In the data collection, we specifically focused on the websites that provide three types of information related to color terms that can be used for the analysis of different aspects of color naming and categorization. Firstly, we obtained linguistic information including the actual color words accompanied in many cases by more specific non-hue features of the color in the product (for more details see section 3.2). Secondly, we gave preference to the websites that give visual representation of the color options in the form of color swatches for specific shades. These visual representations gave us a way to measure and model the referential range of color words and color categories independent of language. Thirdly, we retrieved contextual and sociolinguistic information related to the product, company, target consumer, which can be used for the analysis of constraints these sociolectal factors set on color naming and categorization. The three types of information allow us not only to compare the morpho-syntactic structure of the color terms like *wine*, *wine all mine*, *wine on ice* or *chilled wine*, *sparkling wine*, *wine barrel*, but also to literally ‘see’ and measure how similar these shades are using their RGB coordinates (128, 52, 65; 132, 67, 73; 235, 99, 115; 105, 66, 83; 236, 227, 207; 119, 77, 63). This means that we can explore the complex interactions between color naming and color categorization patterns and further analyze the effects of contextual and sociolectal factors

¹ We would like to especially thank Tom Ruetten whose expertise and help in automatic data retrieval and data processing provided us with additional data making it possible to apply quantitative methods in the study.

on these interactions, since we know that the first three colors are used for makeup (eyeshadows and lipsticks) and are designed for women and the next three are names of premium and budget paints.

However, we need to note the limited linguistic context in the use of color words in this type of material, which makes our analysis different from but complementary with the existing corpus studies. Unlike the traditional text corpora, marketing websites provide hardly any linguistic context for color words, i.e. color words are used as isolated items in their purely nominative function. This type of data has both advantages and disadvantages. On the one hand, the limited linguistic context makes it harder to analyze some linguistic aspects of color naming, for instance their function in a sentence or collocational patterns. On the other hand, the lack of syntactic constraints gives access to the lexical potential and creativity of color words such as *New York state of mind* or *kiss me coral* that can hardly be expected in a text corpus of any length.

2.2 Sampling

Four product categories were selected to be included in the analyses: cars, clothing, makeup, house paints. The selection and sampling procedures were based on two main considerations. On the one hand, we attempted to make the selection as diverse and representative as possible, and include different product categories from as many brands as possible. On the other hand, we had to take into account technical aspects of the data collection, such as machine readability of the data and data availability. For instance, the categories of electronic devices and household equipment originally considered for inclusion were discarded because of the limited visual representation of the color swatches for most products in these categories.

We used two sampling strategies in the data collection. *Retailer sampling* involved automatic extraction of all relevant data from retailer websites that sell products from numerous brands and companies. Specifically we used the following four websites:

- carsdirect.com - the oldest US website selling new and used cars from 60 car brands available in the US market;
- nordstrom.com - the website of the department store chain of the same name that was used for collecting data on makeup from 32 luxury brands;

- saksfifthavenue.com – another department store website that was selectively sampled for clothing from 114 high-end brands;
- clothingwarehouse.com - a retailer website selling clothes from 37 lower-end US brands.

Brand sampling involved manual collection of data from the websites of specific brands such as Avon, Maybelline, JCrew, Banana Republic, Benjamin Moore, Olympic Paints etc.

The two sampling approaches we used are different and complementary in their representativity. The retailer sampling provides data from a wider range of manufacturers but often only on a limited range of products sold by a specific retailer. This approach is more representative of the different brands. Brand sampling is limited to fewer manufacturers due to technical limitations of manual data collection, but it provides a more exhaustive range of products from a specific brand. This approach is more representative of the product range within a brand. The balanced combination of the two sampling strategies allows the advantages of both.

2.3 Method

In the two presented case studies, we applied the statistical method of multiple linear regression, which is a standard statistical technique best suited for measuring the effects of multiple factors on the value of a numeric response variable. Regression methods are based on using a statistical model in the form of a mathematical equation that includes one or more explanatory variables (predictors) whose effect on the studied phenomenon (response variable) we want to explore. A regression model can be seen as a way of reducing the complex reality represented in the data to a number of statistically significant parameters and their interactions whose effect on the studied phenomenon can then be measured and used for predicting the values of the response variable. The advantage of the multiple linear regression is, obviously, that it allows evaluating the effects of multiple predictors included in the model independently from each other as well as their combined effects (for more detail on regression models and their interpretation in linguistic research see Speelman and Geeraerts 2009, Geeraerts and Speelman 2010).

3. Case study 1: variation in the color term length in four product categories

3.1 The explanatory variables

Selecting sociolectal parameters and operationalising them as explanatory variables that could be applied across the four product categories proved to be a major challenge in our analysis that eventually led us to testing two series of linear models of different granularity. To give an example, gender and age of the target consumer are obvious factors to include in the analysis of the clothing domain, where we typically distinguish between collections for women, men, and children, sometimes also for teenagers. However, these consumer-related parameters are much less obvious in cars or house paints, and are biased in makeup, which is predominantly targeted at adult female consumers. In cars, on the other hand, it is common to distinguish between American, Asian and European brands, but this parameter is much less applicable to clothes and is irrelevant in house paints. Furthermore, in several cases the parameters that appear intuitively plausible for all products are in fact confined to individual categories in their values, making non-intersecting classes. For instance, all product categories involve some kind of subcategorization: in cars we can distinguish between passenger cars, SUVs, vans, etc.; in clothes we can speak about shirts, pants, skirts or coats, etc.; in makeup we have lipsticks, foundations and eye shadows, etc. But these subcategories remain limited to a specific product domain in their values and do not apply across all products, which makes them not very good predictors for other categories.

Taking all these considerations into account, in the first case study we have selected three sociolectal parameters to be tested as explanatory variables determining variation in the length of color terms in four product categories: product category, prestige status of the brand, the diversity of color choices in a product.

Product category

The variable ‘product category’ has four values in our sample of 48192 tokens: ‘cars’, ‘clothing’, ‘makeup’, ‘paints’. As has been mentioned, our main objective in choosing the categories was to make the selection as broad and as diverse as possible, and the main constraint was technical

accessibility and availability of the data. The number of tokens for each category is summarized in figure 1.

Our prediction for the impact of product category on the length of the color terms is based on the intuitive observation that the length of a color term and its complexity vary depending on the product category. The longest color terms are found in cars and in makeup, and the shortest in clothes, with paints somewhat in between.

Prestige status of the brand

The variable ‘status’ with its two values ‘premium’ and ‘standard’ captures the intuitive distinction between prestige (luxury, high-end) brands and products, on the one hand, and all the others that are not associated with these characteristics, on the other hand. This distinction can be made in every product category we have included in the analyses, even if to a different degree, which makes it a good candidate for an explanatory variable in the general model measuring the variation in the length of color terms across different products. Our intuitive expectation is that prestige brands might use more elaborate and complex color names that might be longer than those used by non-prestige brands. At the same time, closer observation of the data reveals an alternative tendency in the high-end clothing brands. They, in fact, often use very simple color names, and tend to rely more on the high quality image of the clothing item, rather than on verbose descriptions and color swatches. This is especially obvious when the item comes in only one color, which is often the case with luxury brands that offer designer collections of unique creations rather than mass produced clothes. Multivariate statistical modeling is especially useful in disambiguating this kind of alternative or conflicting tendencies.

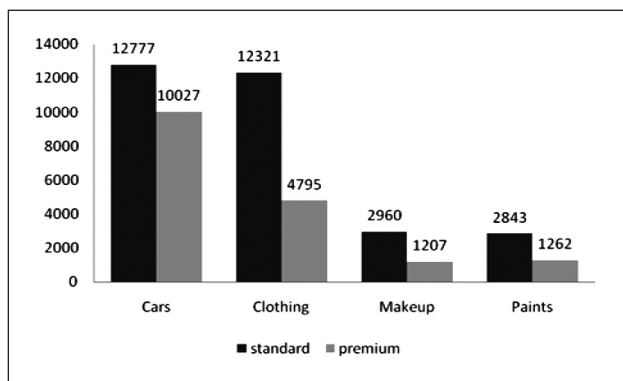
There were some additional difficulties involved in operationalizing and measuring the brand status. The most readily available status-related parameter we have considered was the price of the product. However, several considerations led us to discard price as a direct measurement of prestige status. Firstly, there is a great variation in retail prices for a number of lower-end products in makeup and clothes. For example, we have found that the price of a lip liner by Revlon can range from \$1 to \$30 depending on the retailer or even a specific shade, and blurring the distinction between luxury and budget cosmetics. Secondly, price variation is tied to the product internal subcategorization. For instance, a designer

T-shirt might be close in price to a winter coat from a budget brand, or an inconspicuous van might cost more than a luxury compact car. This made us more cautious about using absolute prices as a direct operationalization of a brand's prestige status. Finally, the concept of prestige or luxury status appears to be tied more closely with the consumer's attitude rather than the absolute price. It is well known that luxury brands often have cheaper entry-level products that allow them to compete with budget brands in price without losing their prestige status in the eyes of the consumer.

Considering all that, we focused on retailer-related parameters and used consumer-oriented resources like consumer guides that often distinguish between luxury and budget brands in a specific product category. For instance, Consumer Guide Auto website systematically distinguishes between premium and non-premium standard car models in their buying guides. This distinction is applied across all types of cars (compact cars, middle size and large cars and SUVs) except for more utilitarian vehicles like pickup trucks and vans, all of which were coded as 'standard'. This information was then generalized as a brand attribute. The brands that mainly produce non-premium models like Chevrolet, Nissan or Ford were coded as 'standard'. The brands whose models are mostly classified as premium were coded as 'premium' brands. This generalization was then verified against native speakers' intuitions. In makeup, we used the lists of luxury and budget makeup brands suggested by shopwiki.com, which were checked against the two partner sites, drugstore.com and beauty.com, that explicitly position themselves as budget and luxury brand retailers respectively. Retailer-related criteria were also used as status indicators in clothing and paints. Clothing brands sold at high-end department stores like Saks Fifth Avenue and Nordstrom, as well as the brands labeled as designer clothes and not sold wholesale (JCrew, Banana Republic), were coded as 'premium'. Brands sold at lower-end shopping centers like Kmart and Sears, or at wholesale websites like clothingwarehouse.com, were coded as having 'standard' status. In paints we also used the internal subdivision of brands within home improvement stores like Lowe's and Home Depot that offer a choice between more expensive premium brands and budget paints.

At the final stage, the coding was verified against the available prices averaged across the brands, and the intuitions of speakers of American English. The distribution of the data across the two status levels in the four product categories is summarized in figure 1.

Figure 1. Distribution of premium and non-premium brands in the four product categories



As a final remark, it should be noted that for the purposes of this analysis we have used the most coarse-grained subdivision into ‘premium’ and ‘standard’ brands, and regarded the premium brands as a marked category associated with special luxury or prestige status. ‘Standard’ brands were treated as a non-marked category including all the brands that do not have particular prestige associations or are regarded as budget products. (This distinction probably requires at least two refinements in further research. Firstly, more fine-grained subdivision of status might better represent the situation in such product categories as cars, clothing, and makeup. Secondly, there might be a within-brand distinction between more expensive luxury products and more affordable entry-level options.)

Diversity of color choices in the product

This parameter was operationalized as a numeric variable ‘number of shades’ and was intended to measure how the additional ‘pressure to name more colors’ might diversify or complicate the structure of color terms. Our expectation based on intuition is that the greater number of shades in a product might demand more diverse and more complex color names. It is reasonable to use a few short and simple terms when there are only a few shades to name, but the task becomes more challenging when you need to name 27 shades of red and pink lipsticks or 53 green paints.

3.2. Response variable

What's in a (color) name?

Working with real language use data creates an additional challenge for analysis, as the retrieved data the researcher has to deal with often do not fit the intuitive expectations very neatly. One of the first obstacles in the automatic processing and analysis of color terms in advertising we have faced was that the lexical units used as color names often contain other (non-color) information.

The additional information included by the marketers into color descriptions can be roughly subdivided into *numeric components*, *pattern words*, and *luster words*.

Numeric descriptions are used in several ways in color names. Firstly, in paints and very often in makeup a numeric code might be systematically added to the color name proper, e.g. *31 dark auburn*, *32 dark mahogany brown*, *33 dark soft brown*, *34 deep burgundy*. In these cases, numbers form a consistent system of oppositions and serve as a parallel identification system that can be used to distinguish all color shades in a specific product. In fact, automotive paints and textile color also have their unique numeric codes in technical specifications, but they are much less present in the purchase situation. Secondly, numbers can have distinctive function within a range of related shades. For instance, in makeup we can find *honey no.1*, *honey no. 2*, *honey no. 3* or *light 1*, *light 2*, *light 3* or even *bg (beige) 911*, *bg 935*, *bg 943*, *bg957*. But in these cases numbers are not added systematically to all products and cannot serve as unique identifiers without the color name. Thirdly, a numeric component can be added to a color name without obvious distinctive function. For instance, in cars we have *black II* and *super red 5* but no other numbers to distinguish from or to suggest a systematic opposition. Finally, numeric components can refer to the number of shades in the product. For instance, in makeup we find *bronze twist duo*, *ice shimmer trio* or *island breeze quad* referring to the number of shades in a compact eye shadow. Similarly, in cars we have *platinum ice tricoat*, describing the number of paint coats.

Pattern and *fabric descriptions* are particularly common in clothing, where the color word can refer to the main color of the clothing item, or include a component to describe the pattern or fabric, as in *peony stripe*, *pewter grey heather*, *charcoal plaid*, *cinnamon tweed*.

Luster and *shine* words are most characteristic of automotive paints and makeup, where they describe the shiny surface color and their different reflectance. We can roughly distinguish four semantic groups of words that are used in this function:

- metals: e.g. in *cream gold*, *blazing copper metallic*, *polished pewter metallic*, *darkest tungsten*, including the most exotic metals in the periodic table, like osmium, palladium, indium, etc.

- minerals and gems: e.g. in *oyster pearl*, *titanium mica*, *willow green opal*, *blue onyx*, etc.

- natural and especially atmospheric phenomena: e.g. in *graphite frost*, *bronze mist*, *starfire* and their derivatives like *gold firemist*, *red firemist metallic*, *steelmist metallic*.

- fabrics: e.g. in *green silk*, *silky beige* or *sterling blue satin glow*.

These ‘non-color’ components in the color terms present difficulties at both the theoretical and the methodological levels. Intuitively, names like *racing green pearl effect*, or *dark green satin clearcoat*, or even *olive silver metallic*, *camel pearl*, *pink promise frost* or *coral kisses creme*, etc. appear to be artificially constructed from a ‘color name proper’ and a luster term like *pearl* and *pearl effect*, *crème* or *frost* that is “asking” to be removed, especially as in many cases we find series of terms distinguished by the luster components alone, like *champagne*, *champagne mist metallic*, *champagne mist pearl*, *champagne pearl*, *champagne pearl clearcoat*, where *champagne* appears to be the ‘real’ color name. However, this intuitive perspective quickly runs into a problem when one is faced with names like *alpine frost* and *arctic frost*, or *inka gold*, *gold leaf* and *sparkling silver*, etc., where the luster referring component is hard to remove. Even names like *parchment gold*, *iridium silver metallic*, *cashmere silver metallic* on a second consideration are not so straightforwardly decomposable. Finally, we find color terms like *satin gold metallic* or *silver pearl metallic* that seem to consist only of luster components.

In other words, there is a continuum of luster components included in color terms to specify the dimension of ‘shininess’ or surface reflectance. This dimension can be compared to the brightness dimension that has been most often discussed in color categorization research among the parameters of color categorization alternative to hue categorization (MacLaury 1992, 1997; Saunders and van Brakel 1997; Lyons 1999; Kertulla 2002). The issue of non-hue aspects of color terms is probably best known from Conklin’s examples of Hanunóo color terms that distinguish wetness vs dryness or

freshness (succulence) vs desiccation rather than hues. Non-hue bases of color categorization have been extensively researched by MacLaury in Meso-American languages, as well as in a number of diachronic studies of English and other languages (Casson 1997, Kertulla 2002). These studies demonstrate that many English color terms developed from brightness words, shifting their meaning to describing hues.

Following a usage-based approach to linguistic research, i.e. exploring language as it is used by the speakers, at this stage of the analysis we retain all the non-hue referencing components in the color names, except for the numeric elements that are used as a parallel system of identification and can be removed without losing the distinction between shades. Our assumption is that we need to start with the most ‘complete’ color names including non-hue referencing components, as they were selected by a language user for this specific situation and type of discourse, in order to give a more complete account of the color naming strategies and speaker choices.

There is also an additional technical consideration in favor of the analysis of complete color words. Identifying non-hue components involves extensive manual coding which often becomes a bottleneck for semantic analysis. Manual data analysis gives more semantically refined results, but at the quantitative scale that we try to operate on, it makes them less applicable for developing automatic procedures and makes the results more vulnerable to the researcher’s subjective judgments.

How long is a color term?

The length of a word can be measured as a continuous variable in a number of different ways. In basic color term studies it is measured in phonemes or syllables (see Corbett and Davies 1997). However, for the purposes of our study these measurement units are not very practical since they would involve extensive manual coding. Instead, we used the measurements standard in the analyses of written corpora: word length measured as a number of lexical components and as a number of characters.

Measuring the length of a color term in lexical components might appear to be a more plausible approach from a linguistic point of view. For example, it is probably more intuitive to think of *very light ash blond* as consisting of four components rather than 17 characters. However, if we take into consideration the automatic language processing demands

of quantitative analysis, measuring the length of a color term in words becomes much less practical. One of the major obstacles is delimiting lexical components in a number of cases.

Firstly, geographical names used in such color terms as *Santa Fe red pearl* or *San Marino red* can be regarded as either one component from a semantic perspective or as two components from the graphical point of view. The decision is even less obvious in cases like *New York grey*, *Pikes Peak white*, *Rocky Mountains light brown*, where the geographical names are more semantically decomposable. Secondly, there is considerable variation in the spelling conventions used by different websites and even within one website. Closer inspection of the data has revealed alternative spellings for words like *rain forest* and *rainforest*, *gun metal* and *gunmetal*, *red fire* and *redfire*, etc. It was even more surprising to see words like *mountaintop* and *horseradish* spelt as one word. For the purposes of our analyses, we followed the original spelling of the color term as it was used by the website. Finally, there is the question of function words and possessive forms. Once we focus on semantic criteria in identifying the meaningful lexical components, the next question to answer is whether components like *from*, *a*, *up*, *'s* in *kiss from a rose*, *spice it up*, or *knight's armor* should be regarded as such.

Alternatively, using the character-version of color term length solves these problems or minimizes their effect. At the same time, the length of the word measured in characters allows more fine-grained measurement and is less distorted by the differences between longer and shorter words.

However, to be on the safe side, we performed a preliminary analysis comparing the linear models for the character-version and the lexical-component version of color term length as a response variable. Both models were constructed using stepwise forward regression. This automatic procedure selects the variables and their interactions in a sequence that significantly improves the statistical model. The models generated for the character-version and the lexical-component version of the color term length were identical in the selected predictors, their sequence and interactions. But at the same time, the adjusted R squared that indicates the unexplained variation in the data was higher for the character-version, which suggests that this model is better suited for describing the variation patterns for color term length in our data. All these considerations led us to opt for the character-version of measuring the response variable in both case studies.

3.3 Results of the statistical analysis

We have applied a multiple linear regression to test the statistical model with three explanatory variables - product category, brand status, and number of shades – and to measure their effects on the length of color terms as a response variable. Before discussing the results of the analysis, we need to make two comments on the diagnostics and initial testing of the model.

Firstly, the analysis of the data distribution has revealed a group of potentially influential observations with an abnormally high number of shades. They came from a paint retailer website, where the palette of 1337 shades was presented to the consumer as one block of color swatches without any subdivision into collections. These observations were removed from the tested dataset in order to avoid the distortions from the unusually high values for the number of shades in the product. This turned out to be a justified solution when the website made available their new redesigned palette, subdivided into more manageable categories of color shades.

Secondly, in the preliminary analysis we have explored the variable interactions selected by the forward stepwise regression to be included in the model. The interaction between the product category and the brand status showed a statistically significant, but very small effect on the value of the response variable. The effect of the interaction between these two variables was limited to 1-2 characters, which we found not practically significant for interpreting the regression output. As a result, we removed the interaction from the model and focused on the model with main effects only in the further analysis. The results of the regression test are summarized in table 1.

Table 1. The results of the multiple linear regression analysis of the color term length in four products

```

-----
Coefficients:
                Estimate   Std.   Error t value
Pr(>|t|)
(Intercept)      7.762963   0.046941  165.377 < 2e-16
***
product.category.auto  9.936364   0.057127  173.935 < 2e-16
***

```



```

product.category.makeup 3.223410    0.099237    32.482    < 2e-16
***
product.category.paint  4.536503    0.144935    31.300    < 2e-16
***
status.premium          -0.518483    0.054621    -9.492    < 2e-16
***
number.shades           -0.010222    0.001623    -6.298    3.04e-10
***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.559 on 46589 degrees of freedom
Multiple R-squared: 0.4039, Adjusted R-squared: 0.4038
F-statistic: 6312 on 5 and 46589 DF, p-value: < 2.2e-16
-----

```

The first column in the table lists the regressors (variables in a specific value) included in the model and the second column shows their estimated effect on the value of the response variable, i.e. the length of the color term. The last column shows the statistical significance of the specific effects. In the presented model the ‘***’ codes show high statistical significance of the effects. The first line (intercept) is the group of observations with reference (or baseline) values for the explanatory variables, i.e. values not listed in the names of the other regressors below. In principle, the reference values for non-numeric variables are selected arbitrarily (in alphabetical order), but for our model we have selected ‘standard clothes’ as the reference group of observations in order to facilitate interpreting the results.

The following lines show the main effects of the variables, i.e. the difference in the predicted length of a color term when the explanatory variable has a value other than the reference value and the effects of all other variables are controlled. By adding the estimate for a specific variable value to the intercept estimate, we can calculate the predicted length of a color term for different groups of observations in our dataset and analyze the effects each explanatory variable included in the model has on the response variable.

Thus the main effects in our model suggest that ‘product category’ is the strongest predictor for the variation in color term length. The category of clothes is characterized by the shortest color terms, with the average color term length approaching 8 characters, which probably suggests that color terms used for clothing tend to be simpler. This effect might

be ‘inherited’ by online advertising from marketing materials designed for brick and mortar clothing stores, where color terms are hardly ever present and might not be expected to affect the purchase decision as much as the color names for makeup or paints. Cars appear to have the longest color terms, with the average length close to 18 characters, which can be attributed to the wider use of non-hue referencing components in the names of automotive paints - mainly luster words like *pearl*, *mica*, *metallic* etc. Makeup and paints take the middle position, with the average length of 11 and 12 characters respectively.

Premium status has a slight negative effect on the response variable, i.e. premium status brands tend to use somewhat shorter color terms. However, it should be noted that a closer analysis of the variable interactions has revealed that the effect of the brand status might be either negative or positive, depending on the product category. This effect will be further explored in the second case study, focusing on one product category, where we can analyze the interaction of the status with other variables in more detail.

Increase in the number of shades also has a slight negative main effect across four product categories, i.e. color terms tend to become shorter when there are more color shades to be named. However, closer inspection of variable interactions has revealed that this effect might vary depending on the product category.

As a final remark we would like to point at one more statistical parameter of the model - adjusted R squared, which indicates how much data variation is explained by the model. In this case it is over 40 per cent, which shows that the model has quite good predictive power.

To sum up the results, the multiple linear regression analysis reveals significant variation in the length of color terms in four product categories. In general, the value of the response variable predicted by the model - length of the color term – is in line with our intuitive prediction based on the data observation, i.e. we find the shortest color names in clothes and the longest ones in cars. The product category itself is the strongest predictor for the response variable, whereas the brand status and diversity of colors in a product have a very slight negative effect, which, however, might be suggestive of different effects depending on the product category.

4. Case study 2: color term length in cars

In the second case study we zoom in on one product category, that of cars. We use the same statistical method of multiple linear regression, but we shift the focus to the product-internal variation in the length of color terms. In this case study we aim to explore additional sociolectal factors specific to a particular product category, and look closer at their interactions with the factors that were suggested in the first case study.

The regression model tested in this case study includes the same response variable operationalizing the length of color terms measured in characters, two explanatory variables analyzed in the previous case study (number of shades in the product and brand status) and three explanatory variables specific for cars as a product category: the country of the parent company, the year of manufacturing, and the type of car. These additional explanatory variables are presented in the next section, followed by a discussion of the results.

4.1 Explanatory variables specific for the product category of cars

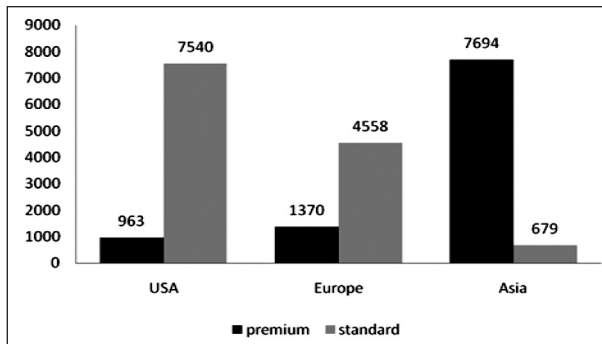
Country of the parent company

This variable operationalizes the intuitive distinction held between American, European and Asian car brands sold in the US market. In the age of transnational corporations, this distinction is becoming more and more blurred when, for instance, SAAB is still perceived as a European brand even though for several years it was owned by GM. Nevertheless, association of a brand with a specific country seems to be quite strongly present in American automotive culture, which was confirmed by our survey among five American native speakers. In the survey, the informants were asked to identify the country which they associate with the specific brands. In the analysis we used a more coarse-grained classification distinguishing between American, Asian and European manufacturers in order to simplify interpreting the regression model. The distribution of tokens in the dataset between different countries and country groups is summarized in table 2, and figure 2 shows the brand subcategorization in the cars from different countries.

Table 2. Distribution of tokens by country of parent company

country	number of tokens	country group	number of tokens
American	8503	America (USA)	8503
British	2600	Europe	8373
French	200		
German	4362		
Italian	587		
Swedish	624		
Japanese	5026	Asia	5928
South Korean	902		

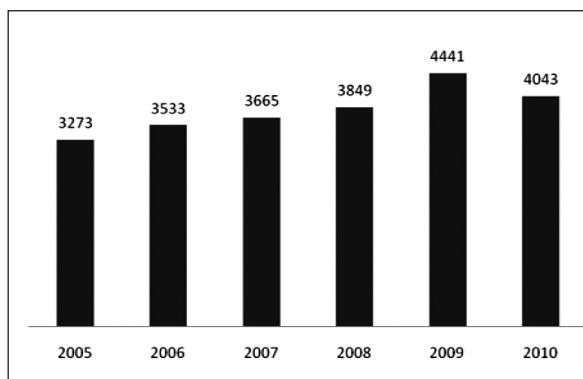
Figure 2. Distribution of premium and non-premium brands in cars



Year of manufacturing and age

These variables are quite straightforward, and are both connected to the year when the car was manufactured and sold. The sample used for the analysis includes observations for the years 2005-2010 as shown in figure 3. After testing a number of models, we also gave preference to a more coarse-grained classification using two categories for the age of cars: 2009-2010 models sold as new cars were coded as such, whereas 2005-2008 models sold as used cars were coded as ‘older’.

Figure 3. Distribution of tokens by year of production



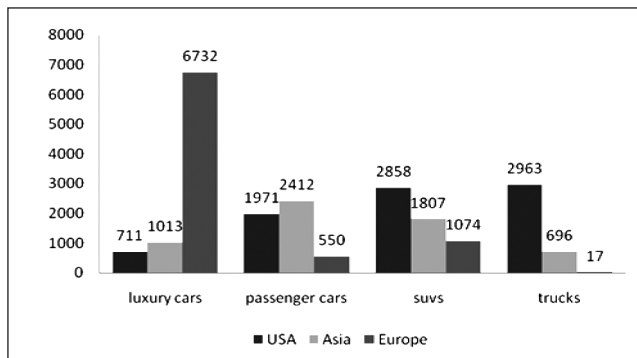
Type of car

This variable was used to operationalize the product-internal subcategorization of cars into different types of vehicles such as passenger cars, SUVs, pickup trucks, vans, sports cars, etc. These categories are quite complex and somewhat arbitrary, involving a number of criteria such as car design, function, price and a number of technical parameters. However, this classification appears to be quite consistently used across car dealer websites, in some cases with further subdivisions. In our sample, we used the most coarse-grained classification into four very broad types: 'luxury cars', 'passenger cars', 'SUVs' and 'trucks' (for the distribution of tokens see figure 4).

Two comments need to be made on the values in this variable. Firstly, the category of 'trucks' includes two types of vehicles - pickup trucks and vans. These two groups were conflated first of all because our dataset includes no tokens for vans manufactured by European brands, which would distort the interactions between the variables in the regression model. At the same time, these two categories are the least numerous and they are qualitatively most similar, namely trucks and vans are more utilitarian vehicles designed primarily to transport cargo and larger groups of people, rather than demonstrate the lifestyle of the owner, for instance. A second remark concerns the category of 'luxury cars'. To a certain degree this category is close in its meaning to the category of premium cars. However, it applies to a different level of granularity. We used the category

of ‘status’ to capture the distinction between brands, whereas the category of ‘luxury cars’ applies to specific models. This way we can capture the distinctions made within brands and test the situation when a brand makes both ‘luxury cars’ and ‘passenger cars’, like Chrysler, whose Chrysler 300 is classified as a luxury car the same as Nissan’s gtr model whereas most models by these manufacturers are classified as non-luxury or ‘standard’. As a result, the brands themselves are not associated with luxury status and were coded as ‘standard’.

Figure 4. Distribution of different types of cars between the brands from different countries



4.2 Results of the statistical analysis

We applied the multiple regression test to the model, including color term length as a response variable and five explanatory variables: country of the parent company, prestige status of the brand, age of the car, type of the car and number of shades for the model. The abbreviations for the variables and their values are summarized in table 3. The values in boldface were selected as reference levels.

Table 3. Variables included in the multiple linear regression model for cars

variable	value	abbreviation
country of the parent company	usa	
	asia	country.asia
	europa	country.europa
prestige status of the brand	standard	
	premium	status.premium
type of car	luxury cars	
	passenger cars	type.passenger.cars
	SUVs	type.suvs
	trucks	type.trucks
age	new	
	older	age.older
number of shades for a model		num.shades

In the preliminary analysis we have applied a forward stepwise regression which selected 8 variables and 8 interaction terms to be included in the model as statistically significant predictors. We have additionally tested the suggested interaction terms with anova and calculated the size of their effect for closer inspection. Although all the interactions showed statistically significant improvement of the model in the anova test, we removed four of them because they were of little practical significance for interpreting the results. The effects showed no cross-over patterns that would suggest a different effect depending on the product category, and their deviations from parallel lines were no more than two or three characters.

The results of the regression analysis of the final model including the main effects and three interaction terms are summarized in table 4.

Table 4. The results of the multiple linear regression analysis of the color term length in cars

```

-----
Coefficients:
value Pr(>|t|)
(Intercept)
54.153 < 2e-16 ***
country.asia
14.280 < 2e-16 ***
country.europe
11.059 < 2e-16 ***
num.shades
-6.916 4.76e-12 ***
status.premium
14.862 < 2e-16 ***
type.passenger.cars
-9.979 < 2e-16 ***
type.suvs
-5.722 1.07e-08 ***
type.trucks
-8.451 < 2e-16 ***
age.older
-3.498 0.000469 ***
country.asia:num.shades
4.561 5.11e-06 ***
country.europe:num.shades
2.947 0.003217 **
country.asia:status.premium
10.062 < 2e-16 ***
country.europe:status.premium
11.114 < 2e-16 ***
status.premium:type.passenger.cars
2.733 0.006286 **
status.premium:type.suvs
7.014 2.37e-12 ***
status.premium:type.trucks
4.871 1.12e-06 ***
num.shades:type.passenger.cars
4.156 3.25e-05 ***
num.shades:type.suvs
-0.175 0.861137

```

	Estimate	Std. Error	t
(Intercept)	26.333796	0.486282	
country.asia	-6.003821	0.420428	-
country.europe	-5.171010	0.467574	-
num.shades	-0.275536	0.039840	
status.premium	-6.230178	0.419211	-
type.passenger.cars	-5.281482	0.529235	
type.suvs	-2.811461	0.491372	
type.trucks	-4.860632	0.575185	
age.older	-0.321916	0.092018	
country.asia:num.shades	0.201685	0.044216	
country.europe:num.shades	0.117260	0.039796	
country.asia:status.premium	3.209118	0.318926	
country.europe:status.premium	4.128487	0.371471	
status.premium:type.passenger.cars	2.882776	1.054910	
status.premium:type.suvs	2.868544	0.408954	
status.premium:type.trucks	5.056644	1.038008	
num.shades:type.passenger.cars	0.186191	0.044803	
num.shades:type.suvs	-0.006767	0.038683	


```

num.shades:type.trucks          0.132176    0.045791
2.886 0.003899 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.634 on 22785 degrees of freedom
Multiple R-squared:  0.09636,    Adjusted R-squared:  0.09564
F-statistic:   135 on 18 and 22785 DF,  p-value: < 2.2e-16
-----

```

The intercept characterizes the baseline group of observations, including new American luxury cars manufactured by non-prestige brands, and they have the average predicted color term length of 23 characters. The next two lines show the main effects of the variable ‘country of the parent company’, which suggests that American manufacturers tend to use somewhat longer names than foreign companies present in the US market.

The variable ‘brand status’ also has quite a strong negative effect. Specifically, the model suggests that premium brands tend to use shorter names for the colors in their models. This effect is much more pronounced in this product-specific model compared to the effect of this variable tested across four product categories. It is possible that the effect of this parameter in the more general model is confounded by the effects of the interaction terms which we did not include in the main effects only model. Below we will discuss the interactions of the brand status with other variables in more detail.

Product-internal subcategorization represented by the variable ‘type of car’ with its four values also has a relatively strong influence on the preferred length of color terms. The longest color terms are found in ‘luxury cars’, which are often associated with style, technological advancement and somewhat excessive prices. From this perspective a more complex color name can be seen as one more feature in the long list of extras offered in luxury cars. The colors of ‘less luxurious’ vehicles such as economy ‘passenger cars’ and ‘trucks’ (including vans) that are expected to have lower prices tend to receive somewhat shorter names. ‘SUVs’ take an intermediate position with color terms only three characters shorter than ‘luxury cars’, which might reflect their ‘more luxurious’ status and higher price, placing them closer to luxury vehicles.

The ‘number of shades’ for a model and its ‘age’ have a very slight negative effect. However, it should be noted that the effect of the numeric

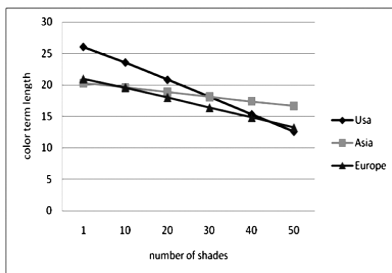
variable ‘number of shades’ is estimated per shade and with the average number of shades close to 13 it can be quite strong. This effect suggests that in the situations where a customer is offered more color choices, color names tend to become shorter, counter to our intuitive prediction. This might be explained by a higher processing load effect, i.e. handling longer names becomes more difficult as they become more numerous.

The effect of the car ‘age’ might also become stronger if we look at even older cars. In this analysis we only included cars sold in 2005-2010.

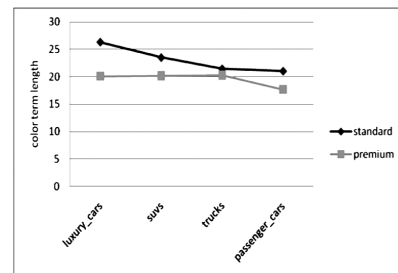
The lower half of the table shows the combined effect of the interactions between variables in the model. The whole idea of interaction terms is that the effect of a specific variable might depend on the value of another explanatory variable and vice versa. In other words, interaction terms can reveal, as has happened in our model, that the effect of a variable can be different depending on the values of another variable. In order to simplify the interpretation of the results, we have only included four pairwise interactions in the model that we found practically significant. Specifically, we discuss below how the country of the parent company and type of car interact with the brand status and number of shades for a specific model. The analysis is based on the calculated effects of the variable interactions that are represented in the plots (figure 5). The interactions are plotted in the sequence suggested by forward stepwise regression and indicative of their statistical significance.

Figure 5. Interactions between variables in the linear regression model for cars

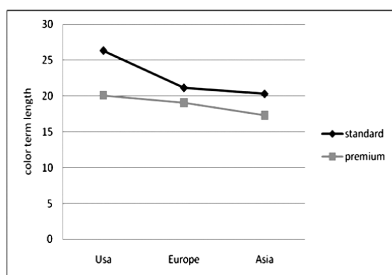
a) ‘country of the parent company’ and ‘number of shades’



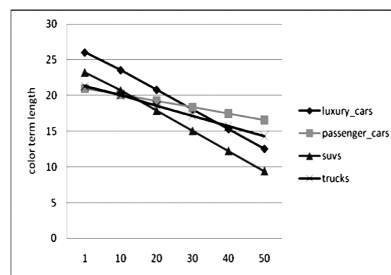
c) ‘car type’ and ‘brand status’



b) 'country of the parent company' and
'brand status'



d) 'car type' and 'number of shades'



Plot a) and plot b) show how the length of the color term changes relative to the 'number of shades' depending on the 'country of the parent company' or 'type of car'. Both plots are similar in the general downward tendency in line with the main negative effect of the variable 'number of shades'. However, closer inspection of the interactions reveals that different categories of cars are sensitive to this tendency to a different degree.

American cars show the steepest slope indicating more variation in the length of color names, i.e. color names used by American manufacturers become more than two times shorter with the increase in the number of shades. Asian cars, on the contrary, have the most gradual slope indicating their higher consistency in the preferred color term length, which only shorten by four characters. European cars take the middle position in the steepness of the slope. What is interesting is that European models with fewer colors appear to be similar to Asian cars, whereas models with more color options behave in a way similar to American ones.

Different types of cars show even more sensitivity to the increase in the 'number of shades'. 'Luxury cars' and 'SUVs' are characterized by the highest variation in the preferred color term length, indicated by the steeper slopes. Starting from the longest color terms they drop to more than twice shorter length. This effect is in line with the two main effects we have discussed above. Specifically, these two types of cars associated with more diversity in all kinds of detail receive more complex color names, but at the same time with the increase in the number of shades, these extra-long names become too cumbersome to handle or get lost in the growing number of special designer features. On the other hand, 'passenger cars'

and 'trucks' (+vans) once again appear more similar in their behavior and are more consistent in the preferred length of color terms. In fact, color names for 'passenger cars' only change by four characters, which does not make up even one word. With the most gradual slope 'passenger cars' create an interesting crossover pattern. Starting with the shortest color names in the models with fewer color options, they end up with essentially the same color names, which become longer than the color names used for 'luxury cars' and 'SUVs' with similarly diverse color choices.

The interactions of the 'brand status' with the different 'country groups' (plot b) and 'car types' (plot c) also have a similar general pattern, with the 'premium' brands in general using shorter color terms than non-premium brands. However, this distinction appears to be more relevant for some car categories than for others. Thus American 'premium' and 'standard' brands are apparently more distinct in the preferred color term length than foreign brands. This tendency probably reflects not only specific marketing strategies employed by premium and non-premium American car makers, but also the limited presence of non-premium foreign brands, especially European, in the US car market. Plot c) also reveals a more pronounced distinction between premium and non-premium brands in the 'luxury car' sector. This suggests that non-premium brands tend to use longer names even when they make luxury models. In 'passenger cars' and 'SUVs' the distinction between premium and non-premium brands is smaller, though the latter are consistently characterized by somewhat longer color names,. What is more interesting, in 'trucks' (+vans) this distinction becomes irrelevant. This can be attributed to the fact that few premium brands manufacture these car types, though the situation is changing in the pickup-truck sector.

To summarize the discussion, multiple linear regression analysis reveals significant variation in the preferred color term length in the category of cars. The test results indicate that color term length is determined both by more general factors potentially applicable to other product categories (like 'brand status') as well as product-specific factors (like 'type of car' or 'year of manufacturing'). The complex interactions between the variables suggest that these factors can have different effects depending on the group of cars.

5. Conclusions and future perspectives

In the two presented case studies we have explored the variation in color term length relative to several sociolectal factors relevant for the use of color terms in online marketing and sales. Unlike the mainstream basic color categorization studies that address typological or cross-cultural variation using the length of color terms correlated with frequency as a formal parameter of basicness, we focus on the language-internal variation in color naming. We view color term length as a formal feature that might indicate the semantic complexity of the color terms and that varies depending on the usage situation. In other words, we take a sociolinguistic context-based perspective on color naming and categorization.

In the first case study we looked at color names used in four product categories (cars, clothes, makeup, paints) and identified more general sociolectal factors that influence the length of a color term. In the second case study, we focused on the product category of cars, exploring the effects of product-specific parameters and their interaction with more general sociolectal factors.

The specific predictions we have tested can be summarized as follows:

1) There is significant variation in the length of color terms across four product categories.

2) The status of the brand as a sociolectal parameter influences the length of the color term and its complexity. Data inspection, in fact, shows conflicting tendencies with high-end clothing brands using simpler color names, where we could expect more complexity.

3) There is significant variation in color term length within one product category reflecting the product-internal subdivisions (for instance different types of cars or different countries associated with the brands).

4) Diversity of colors in a specific product measured as 'number of shades' might call for the use of more complex names to distinguish them.

Predictions 1 and 2 were tested in the first case study. The analysis confirmed prediction 1, revealing significant variation between four product categories included in the model. The shortest color terms were found in clothes and the longest in cars, in line with our expectations based on data inspection. Product category also proved to be the strongest explanatory variable for the studied variation. Furthermore, if we compare the amount

of explained variation in the two models indicated by adjusted R squared, we can conclude that the four-product model including 'product category' as a predictor explains a larger proportion of variation in the data than the model limited to one product.

Regarding prediction 2, both case studies have revealed a negative effect of the premium status on the length of the color term, suggesting that premium brands tend to use shorter color names. This result contradicted our intuitive assumption, but confirmed the tendency we observed in the data.

Predictions 3 and 4 were tested in the second case study. For prediction 3, the analysis revealed significant variation in the preferred color term length between different subcategories of cars, reflecting the distinction in marketing strategies developed for different types of cars and by brands from different countries. The highest variation was found in luxury cars manufactured by US brands. The close inspection of interactions between the variables has shown that the effects of the studied factors can vary for different groups of observations. In some cases, the effect of a predictor can be neutralized for a specific category, for instance the distinction between premium and non-premium brands in foreign cars and in the category of trucks.

Contrary to our prediction 4, a larger number of color shades in a product has a negative effect on the length and complexity of color terms. In other words, color names become shorter when they come in larger quantities. This effect also varies depending on the group of observations. For instance, American cars are more sensitive to the increase in the number of shades, the same as luxury cars and SUVs, whereas Asian cars and passenger cars are more consistent in the preferred color term length.

The positive results, revealing the significant effects of sociolectal factors on the length of color terms, make us optimistic about the future development of the study. As the next steps we see the following extensions:

- refining the coding schema for the sociolectal factors, including further product-specific and product-independent factors for more fine-grained analysis;
- exploring the effects of sociolectal factors on other formal and semantic aspects of linguistic variation in color naming strategies;
- disentangling the interactions between linguistic, sociolectal, and conceptual factors involved in the variation of color terms.

References

- Bergh, G. 2007. The semiosis of Swedish car color names: Descriptive and amplifying functions. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.), *Anthropology of Color: Interdisciplinary Multilevel Modeling*. Amsterdam: John Benjamins, pp. 337-345.
- Casson, R.W. 1997. Color shift: Evolution of English color terms from brightness to hue. In Hardin, C.L., Maffi, L. (eds.), *Colour Categories in Thought and Language*. Cambridge: Cambridge University Press, pp. 224-239.
- Cencig, E. 1990. United Colours: Colour terms in contemporary fashion trends. *Fachsprache*, 12. 3-4, pp. 114-128.
- Corbett, G. and I. R. L. Davies. 1997. Establishing Basic Color Terms: Measures and Techniques. In Hardin, C.L., Maffi, L. (eds.), *Colour Categories in Thought and Language*. Cambridge: Cambridge University Press, pp. 197-223.
- Geeraerts, D. 2010. *Theories of Lexical Semantics*. Cambridge: Cambridge University Press.
- Geeraerts, D. and D. Speelman. 2010. Heterodox concept features and onomasiological heterogeneity in dialects. In Geeraerts, D., G. Kristiansen, Y. Peirsman (eds.), *Advances in Cognitive Sociolinguistics*. Berlin, New York: Mouton de Gruyter, pp. 23-41.
- Geeraerts, D., S. Grondolaers and P. Bakema. 1994. *The Structure of Lexical Variation. Meaning, Naming, and Context*. Berlin, New York: Mouton de Gruyter.
- Geeraerts, D., G. Kristiansen and Y. Peirsman (eds.). 2010. *Advances in Cognitive Sociolinguistics*. Berlin, New York: Mouton de Gruyter.
- Graumann, A. 2007. Color names and dynamic imagery. In Plümacher, M., Holz, P. (eds.), *Speaking of Colors and Odors*. Amsterdam: John Benjamins, pp. 129-140.
- Jameson, K.A. and N. Alvarado. 2003. The relational correspondence between category exemplars and names. *Philosophical Psychology*, 16, pp. 23-47.
- Kerttula, S. 2002. *English Colour Terms: Etymology, Chronology, and Relative Basicness*. Mémoires de la Société Néophilologique de Helsinki, Vol. LX. Helsinki: Société Néophilologique.

- Klaus, H. 1989. Beobachtungen zu den Modefarbenwörtern in der deutschen Gegenwartssprache. *Zeitschrift für Germanistische Linguistik*, 17, pp. 22-57.
- Kristiansen, G. and R. Dirven (eds.). 2008. *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*. Berlin, New York: Mouton de Gruyter.
- Lyons, J. 1999. The vocabulary of color with particular reference to ancient Greek and classical Latin. In Borg, A. (ed.), *The Language of Color in the Mediterranean*. Stockholm: Almqvist & Wiksell International, pp. 38–75.
- MacLaury, R.E. 1992. From brightness to hue: An explanatory model of color-category evolution. *Current Anthropology*, 11:2, pp. 137-186.
- MacLaury, R.E. 1997. *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. Austin: University of Texas Press.
- Moroney, N. 2008. *The Color Thesaurus*. Hewlett-Packard Company, Magcloud.com digital publication.
- Plümacher, M. 2007. Color perception, color description and metaphor. In Plümacher, M., Holz, P. (eds.), *Speaking of Colors and Odors*. Amsterdam: John Benjamins, pp. 61-84.
- Saunders, B.A.C. and J. van Brakel. 1997. Are there non-trivial constraints on colour categorization? *Behavioral and Brain Sciences*, 20, pp. 167-228.
- Speelman, D. and D. Geeraerts. 2009. Causes for causatives: The case of Dutch *doen* and *laten*. In Sanders, T., Sweetser, E. (eds.), *Causal Categories in Discourse and Cognition*. Berlin: Mouton de Gruyter, pp. 173-203.
- Steinvall, A. 2002. *English Color Terms in Context*. Ph.D. Dissertation. Umeå University: Skrifter från moderna språk 3.
- Stoeva-Holm, D. 1996. *Farbbezeichnungen in deutschen Modetexten. Eine morphologisch-semantische Untersuchung*. Diss. [Studia Germanistica Upsaliensia, 34]. Uppsala: Universitetsförlag Almqvist & Wiksell.
- Stoeva-Holm, D. 2007. Color terms in fashion. In MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.), *Anthropology of Color: Interdisciplinary Multilevel Modeling*. Amsterdam: John Benjamins, pp. 421-440.
- Wierzbicka, A. 1990. The meaning of color terms: semantics, culture, and cognition. *Cognitive Linguistics*, 1-1, pp. 99-150.
- Wyler, S. 2007. Color terms between elegance and beauty: The verbalization of color with textiles and cosmetics. In Plümacher, M., Holz, P. (eds.),

Speaking of Colors and Odors. Amsterdam: John Benjamins, pp. 113-128.

Accepted for publication on 1 October 2010.

Алена Анишчанка, Дирк Спелман, Дирк Херартс
Универзитет у Лувену, Белгија

СОЦИОЛЕКАТСКЕ ВАРИЈАЦИЈЕ У ДУЖИНИ НАЗИВА ЗА БОЈЕ У РЕКЛАМНИМ ОГЛАСИМА

Сажетак

У раду се испитује дужина назива за боје који се користе у рекламним огласима као један од формалних параметара лексичке сложености тих назива. Представљени су резултати два истраживања у којима се применом методе вишеструке линеарне регресије тестирају варијације у дужини назива за боје у односу на два скупа социолингвистичких параметара који су релевантни за оглашавање и куповину преко интернета. Прво истраживање открива најопштије обрасце варијација за четири категорије производа (аутомобили, одећа, женска шминка, боје за кућу) под дејством социолингвистичких фактора као што су категорија производа и статус робне марке. У другом истраживању излаже се детаљније нијансирана анализа једне категорије производа (аутомобила), у којој се фокусирају социолингвистички параметри специфични за дати производ (тип аутомобила, година производње, земља порекла матичне компаније) и њихова интеракција са општијим варијаблама установљеним у првом истраживању.

Кључне речи: дужина назива за боје, социолекатска варијација, категоризација боја, оглашавање, метод вишеструке линеарне регресије

